

HW



PCT09

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/980,771A

DATE: 04/11/2002

TIME: 12:45:14

Input Set : A:\EP.txt

Output Set: N:\CRF3\04112002\I980771A.raw

3 <110> APPLICANT: CNRS
5 <120> TITLE OF INVENTION: STARCH GRANULES CONTAINING A RECOMBINANT POLYPEPTIDE OF
INTEREST, A
6 METHOD OF OBTAINING THEM AND THEIR USES
8 <130> FILE REFERENCE: WOB 99 AB CNR AMYL
10 <140> CURRENT APPLICATION NUMBER: US 09/980,771A
11 <141> CURRENT FILING DATE: 2001-11-15
13 <150> PRIOR APPLICATION NUMBER: FR 99/06494
14 <151> PRIOR FILING DATE: 1999-05-21
16 <160> NUMBER OF SEQ ID NOS: 9
18 <170> SOFTWARE: PatentIn version 3.1
20 <210> SEQ ID NO: 1
21 <211> LENGTH: 3117
22 <212> TYPE: DNA
23 <213> ORGANISM: Chlamydomonas reinhardtii
25 <400> SEQUENCE: 1
26 gccaggcgga gcgtatggct gttgcctcta ccagccgccc cagcagcgcg cgtcctatcg 60
28 tcatcaatgc cgcgtcgttc ggtgtcaaga agaccgcgaa ccagctgctg cgtgagcttg 120
30 ctctgtggctc cgcacgcaag tccacctcgc gctcggctgt tactgggtgcc actgggtgcca 180
32 ctctgcgcgct ggacatcgtg atggttgcgt ctgaggtcgc cccttggtcc aagacgggcg 240
34 gcctggggcga tgtgactggt ggctgccta ttgagctggt caagcgcggc caccgcgtca 300
36 tgaccattgc ccctcgtac gaccagtacg ctgacgcctg ggacacctcg gtggtcgttg 360
38 acatcatggg cgagaaggte cgctacttcc actccatcaa gaagggcggt caccgcgtgt 420
40 ggattgacca cccctgggtc ctggccaagg tctggggcaa gaccggctcc aagctgtacg 480
42 gccccgcgtc cggcgctgac tacctggaca accacaagcg ctctgccttg ttctgcaagg 540
44 ccgctattga ggctgcccgc gtgctgccct tcggcccccg cgaggactgc gtcttcgttg 600
46 ccaacgactg gcactccgcc ctgggtgccg tcctgctgaa ggacgagtac cagcccaagg 660
48 gccagttcac caaggccaag tcggtgctgg ctatccacaa catgccttc cagggccgca 720
50 tgtgggagga ggctttcaag gacacgaagc tgccccaggc cgcctttgac aagctggcct 780
52 tctcggacgg ctatgccaa gtttacactg aggccacccc catggaggag gacgagaagc 840
54 ccccgctgac gggaaagacc tacaagaaga tcaactggct gaaggggtggc attatcgccg 900
56 ccgacaagct ggtgactgtg tcgcccact acgcgaccga gatcgctgcc gatgcgcgcg 960
58 gcggtgtgga gctggacacc gtcacccgcg ccaagggcat tgagggcatt gtgaacggca 1020
60 tggacattga ggagtggaa cccaagaccg acaagttcct gtctgcgccc tacgaccaga 1080
62 acagcgtcta cgcgggcaag gccgcgcca aggaggccct gcaggccgag ctgggcctgc 1140
64 ctgtggaccc caccgcccc ctgttcgcct tcatcgccg cctggaggag cagaagggtg 1200
66 tggacatcat cctggccgcc ctgcccaga tcctggccac cccaagggtg cagatcgcca 1260
68 tcctgggtac cggcaaggcc gcctacgaga agctggtgaa cgccatcggc accaagtaca 1320
70 agggccgcgc caagggcggt gtcaagttct cggcgcccct ggcgacatg ctcaccgcgc 1380
72 gcgcccactt catgctggtg ccctcgcgt tcgagccctg cggcctgac cagctgcacg 1440
74 ccatgcacta cggtagcgtg cccgtggtag cctccaccgg cggcctggtc gacacgtca 1500
76 aggaggcgct caccggcttc cacatgggcg cctgaaccc cgacaagctg gacgaggctg 1560
78 acgcccgcgc cctggccgcc accgtgcgcc gtgccagcga ggtgtttgcg ggcggccgct 1620
80 accccgagat ggtggccaac tgcacagcc aggacctgtc ctggtccaag cccgcccaga 1680

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/980,771A

DATE: 04/11/2002

TIME: 12:45:14

Input Set : A:\EP.txt

Output Set: N:\CRF3\04112002\I980771A.raw

```

82 agtgggaggg cctgctggag gaggtggtgt acggcaaggg cggcgtggcc accgccaaga 1740
84 aggaggagat caaggtgccc gttgccgaga agatccccgg cgacctgcc gccgtgtcct 1800
86 acgcccccaa caccctgaag cccgtgtccg cctccgtgga gggcaacggc gccgcccgcg 1860
88 ccaaggtcgg caccaccgcc cccgccatgg gcgcgtggcg cgcgaccacc cctcggggcc 1920
90 cctcgcccgc cgccgccacc cccaaggtga ccacctacaa gcccgcctg cccgccaccg 1980
92 ccaagcccaa gaccgctggc ctcaagctgg ccggtgaggg ctccaccacc tcgacctcgg 2040
94 agaacggcgc tgcctccaac ggcaacggca acggtgcctc ggctccaag acctcggctg 2100
96 ccaagcccct ggtctccgcc gccaccgcga agtcgccta aagcggcagt agccgcagag 2160
98 gcggcgacag catgagcggc tcgaccaaag ctgtggcagg aacggctgta gcagcggcag 2220
100 gcggccgcca ccggcgagga gcaggcttgc ggcagcgagg gcgatgagct tagcggccgt 2280
102 gagcatggca ggcggaacg tgtgtactga aatgtggtgc atgagagtgt cgtgctgtaa 2340
104 tgaagtcggg tttgcgagac cggagaaacg ccggtttggt tttgtagtgc agggcctgtg 2400
106 gtttcggttt tgcccaagtc caaaagaaga gtaacgaaac tgtagcagta gcagagcact 2460
108 tgcgcggcgc ggcgaccacg ccggcccgtg cgcagcctgt cctgccctca gccttgtgat 2520
110 tcggcggcaa gagggcgggt ctgtacactc catccattcc aggatttttg caggctgcct 2580
112 gagagtttgc cattttgtgg gacgtgagcg gcgggacggc cgcgcccggc tctcctaccg 2640
114 cctccggcaa cggagaagtg ggaggcgctg tagcccggtg acccccgaat gtagaggatg 2700
116 ggatacataa gacgtgtggg aatggtggtg aaagaggagg ggcctgggtc gccctcgat 2760
118 ggttttgttg aggtgcagac ggcaccgtcg gcgtcaaagg ccctcgcaag gcccggtgc 2820
120 cttgggctca ttttgggtgc ccgtcgatga tgagagattg gccagcgggt ttttgaggct 2880
122 ggctcgaagc gagggtttgt ggaagtggag cgaggagggt tggagaaaga ggcgacatg 2940
124 cttgactgga ggtacacaaa gtggagcgtg cgacggcacg gaggcattgg cggactattg 3000
126 acccagtagt gtgaaagta gttggacctg aattctttga gagtaccgcg cattaatccg 3060
128 tgagagagta acaaagatgg cacctgaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa 3117

```

131 <210> SEQ ID NO: 2

132 <211> LENGTH: 2124

133 <212> TYPE: DNA

134 <213> ORGANISM: artificial sequence

136 <220> FEATURE:

137 <223> OTHER INFORMATION: fragment of the complete sequence of cDNA coding for the
GBSSI of

138 Chlamydomonas reinhardtii

140 <220> FEATURE:

141 <221> NAME/KEY: CDS

142 <222> LOCATION: (1)..(2124)

143 <223> OTHER INFORMATION:

146 <400> SEQUENCE: 2

```

147 atg gct gtt gcc tct acc agc cgc ccc agc agc gcg cgt cct atc gtc 48
148 Met Ala Val Ala Ser Thr Ser Arg Pro Ser Ser Ala Arg Pro Ile Val
149 1 5 10 15
151 atc aat gcc gcg tcg ttc ggt gtc aag aag acc gcg aac cag ctg ctg 96
152 Ile Asn Ala Ala Ser Phe Gly Val Lys Lys Thr Ala Asn Gln Leu Leu
153 20 25 30
155 cgt gag ctt gct cgt ggc tcc gca cgc aag tcc acc tcg cgc tcg gct 144
156 Arg Glu Leu Ala Arg Gly Ser Ala Arg Lys Ser Thr Ser Arg Ser Ala
157 35 40 45
159 gtt act ggt gcc act ggt gcc act tgc gcg ctg gac atc gtg atg gtt 192
160 Val Thr Gly Ala Thr Gly Ala Thr Cys Ala Leu Asp Ile Val Met Val
161 50 55 60
163 gct gct gag gtc gcc cct tgg tcc aag acg ggc ggc ctg ggc gat gtg 240

```

RAW SEQUENCE LISTING

DATE: 04/11/2002

PATENT APPLICATION: US/09/980,771A

TIME: 12:45:14

Input Set : A:\EP.txt

Output Set: N:\CRF3\04112002\I980771A.raw

164	Ala	Ala	Glu	Val	Ala	Pro	Trp	Ser	Lys	Thr	Gly	Gly	Leu	Gly	Asp	Val	
165	65					70					75					80	
167	act	ggt	ggc	ctg	cct	att	gag	ctg	gtc	aag	cgc	ggc	cac	cgc	gtc	atg	288
168	Thr	Gly	Gly	Leu	Pro	Ile	Glu	Leu	Val	Lys	Arg	Gly	His	Arg	Val	Met	
169					85					90					95		
171	acc	att	gcc	cct	cgc	tac	gac	cag	tac	gct	gac	gcc	tgg	gac	acc	tcg	336
172	Thr	Ile	Ala	Pro	Arg	Tyr	Asp	Gln	Tyr	Ala	Asp	Ala	Trp	Asp	Thr	Ser	
173				100					105					110			
175	gtg	gtc	gtg	gac	atc	atg	ggc	gag	aag	gtc	cgc	tac	ttc	cac	tcc	atc	384
176	Val	Val	Val	Asp	Ile	Met	Gly	Glu	Lys	Val	Arg	Tyr	Phe	His	Ser	Ile	
177			115					120					125				
179	aag	aag	ggc	gtg	cac	cgc	gtg	tgg	att	gac	cac	ccc	tgg	ttc	ctg	gcc	432
180	Lys	Lys	Gly	Val	His	Arg	Val	Trp	Ile	Asp	His	Pro	Trp	Phe	Leu	Ala	
181		130				135					140						
183	aag	gtc	tgg	ggc	aag	acc	ggc	tcc	aag	ctg	tac	ggc	ccc	cgc	tcc	ggc	480
184	Lys	Val	Trp	Gly	Lys	Thr	Gly	Ser	Lys	Leu	Tyr	Gly	Pro	Arg	Ser	Gly	
185	145				150					155					160		
187	gct	gac	tac	ctg	gac	aac	cac	aag	cgc	ttc	gcc	ctg	ttc	tgc	aag	gcc	528
188	Ala	Asp	Tyr	Leu	Asp	Asn	His	Lys	Arg	Phe	Ala	Leu	Phe	Cys	Lys	Ala	
189				165					170					175			
191	gct	att	gag	gct	gcc	cgc	gtg	ctg	ccc	ttc	ggc	ccc	ggc	gag	gac	tgc	576
192	Ala	Ile	Glu	Ala	Ala	Arg	Val	Leu	Pro	Phe	Gly	Pro	Gly	Glu	Asp	Cys	
193			180					185					190				
195	gtc	ttc	gtg	gcc	aac	gac	tgg	cac	tcc	gcc	ctg	gtg	ccc	gtc	ctg	ctg	624
196	Val	Phe	Val	Ala	Asn	Asp	Trp	His	Ser	Ala	Leu	Val	Pro	Val	Leu	Leu	
197			195				200					205					
199	aag	gac	gag	tac	cag	ccc	aag	ggc	cag	ttc	acc	aag	gcc	aag	tcg	gtg	672
200	Lys	Asp	Glu	Tyr	Gln	Pro	Lys	Gly	Gln	Phe	Thr	Lys	Ala	Lys	Ser	Val	
201		210				215					220						
203	ctg	gct	atc	cac	aac	atc	gcc	ttc	cag	ggc	cgc	atg	tgg	gag	gag	gct	720
204	Leu	Ala	Ile	His	Asn	Ile	Ala	Phe	Gln	Gly	Arg	Met	Trp	Glu	Glu	Ala	
205	225				230					235					240		
207	ttc	aag	gac	acg	aag	ctg	ccc	cag	gcc	gcc	ttt	gac	aag	ctg	gcc	ttc	768
208	Phe	Lys	Asp	Thr	Lys	Leu	Pro	Gln	Ala	Ala	Phe	Asp	Lys	Leu	Ala	Phe	
209				245					250					255			
211	tcg	gac	ggc	tat	gcc	aag	gtt	tac	act	gag	gcc	acc	ccc	atg	gag	gag	816
212	Ser	Asp	Gly	Tyr	Ala	Lys	Val	Tyr	Thr	Glu	Ala	Thr	Pro	Met	Glu	Glu	
213			260					265					270				
215	gac	gag	aag	ccc	ccg	ctg	acg	gga	aag	acc	tac	aag	aag	atc	aac	tgg	864
216	Asp	Glu	Lys	Pro	Pro	Leu	Thr	Gly	Lys	Thr	Tyr	Lys	Lys	Ile	Asn	Trp	
217		275					280				285						
219	ctg	aag	ggt	ggc	att	atc	gcc	gcc	gac	aag	ctg	gtg	act	gtg	tcg	ccc	912
220	Leu	Lys	Gly	Gly	Ile	Ile	Ala	Ala	Asp	Lys	Leu	Val	Thr	Val	Ser	Pro	
221		290				295					300						
223	aac	tac	gcg	acc	gag	atc	gct	gcc	gat	gcc	gcc	ggc	ggt	gtg	gag	ctg	960
224	Asn	Tyr	Ala	Thr	Glu	Ile	Ala	Ala	Asp	Ala	Ala	Gly	Gly	Val	Glu	Leu	
225	305				310					315					320		
227	gac	acc	gtc	atc	cgc	gcc	aag	ggc	att	gag	ggc	att	gtg	aac	ggc	atg	1008
228	Asp	Thr	Val	Ile	Arg	Ala	Lys	Gly	Ile	Glu	Gly	Ile	Val	Asn	Gly	Met	

RAW SEQUENCE LISTING

DATE: 04/11/2002

PATENT APPLICATION: US/09/980,771A

TIME: 12:45:14

Input Set : A:\EP.txt

Output Set: N:\CRF3\04112002\I980771A.raw

229		325		330		335		
231	gac att gag	gag tgg aac ccc aag acc	gac aag ttc ctg tct gcg ccc	1056				
232	Asp Ile Glu	Glu Trp Asn Pro Lys Thr	Asp Lys Phe Leu Ser Ala Pro					
233		340		345		350		
235	tac gac cag	aac agc gtc tac gcc ggc aag gcc gcc	gcc aag gag gcc	1104				
236	Tyr Asp Gln	Asn Ser Val Tyr Ala Gly Lys Ala Ala	Ala Lys Glu Ala					
237		355		360		365		
239	ctg cag gcc	gag ctg ggc ctg cct gtg gac ccc acc gcc ccc	ctg ttc	1152				
240	Leu Gln Ala	Glu Leu Gly Leu Pro Val Asp Pro Thr	Ala Pro Leu Phe					
241		370		375		380		
243	gcc ttc atc	ggc cgc ctg gag gag cag aag ggt gtg gac	atc atc ctg	1200				
244	Ala Phe Ile	Gly Arg Leu Glu Glu Gln Lys Gly Val	Asp Ile Ile Leu					
245	385		390		395		400	
247	gcc gcc ctg	ccc aag atc ctg gcc acc ccc aag gtg cag	atc gcc atc	1248				
248	Ala Ala Leu	Pro Lys Ile Leu Ala Thr Pro Lys Val	Gln Ile Ala Ile					
249		405		410		415		
251	ctg ggt acc	ggc aag gcc gcc tac gag aag ctg gtg aac	gcc atc ggc	1296				
252	Leu Gly Thr	Gly Lys Ala Ala Tyr Glu Lys Leu Val	Asn Ala Ile Gly					
253		420		425		430		
255	acc aag tac	aag ggc cgc gcc aag ggc gtg gtc aag ttc	tcg gcg ccc	1344				
256	Thr Lys Tyr	Lys Gly Arg Ala Lys Gly Val Val Lys	Phe Ser Ala Pro					
257		435		440		445		
259	ctg gcg cac	atg ctc acc gcc ggc gcc gac ttc atg	ctg gtg ccc tcg	1392				
260	Leu Ala His	Met Leu Thr Ala Gly Ala Asp Phe Met	Leu Val Pro Ser					
261		450		455		460		
263	cgc ttc gag	ccc tgc ggc ctg atc cag ctg cac gcc	atg cac tac ggt	1440				
264	Arg Phe Glu	Pro Cys Gly Leu Ile Gln Leu His Ala	Met His Tyr Gly					
265	465		470		475		480	
267	acc gtg ccc	gtg gta gcc tcc acc ggc ggc ctg gtc	gac acc gtc aag	1488				
268	Thr Val Pro	Val Val Ala Ser Thr Gly Gly Leu Val	Asp Thr Val Lys					
269		485		490		495		
271	gag ggc gtc	acc ggc ttc cac atg ggc gcc ctg aac	ccc gac aag ctg	1536				
272	Glu Gly Val	Thr Gly Phe His Met Gly Ala Leu Asn	Pro Asp Lys Leu					
273		500		505		510		
275	gac gag gct	gac gcc gac gcc ctg gcc gcc acc	gtg cgc cgt gcc agc	1584				
276	Asp Glu Ala	Asp Ala Asp Ala Leu Ala Ala Thr	Val Arg Arg Ala Ser					
277		515		520		525		
279	gag gtg ttt	gcg ggc ggc cgc tac ccc gag atg	gtg gcc aac tgc atc	1632				
280	Glu Val Phe	Ala Gly Gly Arg Tyr Pro Glu Met	Val Ala Asn Cys Ile					
281		530		535		540		
283	agc cag gac	ctg tcc tgg tcc aag ccc gcc cag	aag tgg gag ggc ctg	1680				
284	Ser Gln Asp	Leu Ser Trp Ser Lys Pro Ala Gln	Lys Trp Glu Gly Leu					
285	545		550		555		560	
287	ctg gag gag	gtg gtg tac ggc aag ggc ggc gtg	gcc acc gcc aag aag	1728				
288	Leu Glu Glu	Val Val Tyr Gly Lys Gly Val Ala	Thr Ala Lys Lys					
289		565		570		575		
291	gag gag atc	aag gtg ccc gtt gcc gag aag atc	ccc ggc gac ctg ccc	1776				
292	Glu Glu Ile	Lys Val Pro Val Ala Glu Lys Ile	Pro Gly Asp Leu Pro					
293		580		585		590		

RAW SEQUENCE LISTING

DATE: 04/11/2002

PATENT APPLICATION: US/09/980,771A

TIME: 12:45:14

Input Set : A:\EP.txt

Output Set: N:\CRF3\04112002\I980771A.raw

```

295 gcc gtg tcc tac gcc ccc aac acc ctg aag ccc gtg tcc gcc tcc gtg      1824
296 Ala Val Ser Tyr Ala Pro Asn Thr Leu Lys Pro Val Ser Ala Ser Val
297          595          600          605
299 gag ggc aac ggc gcc gcc gcg ccc aag gtc ggc acc acc gcc ccc gcc      1872
300 Glu Gly Asn Gly Ala Ala Ala Pro Lys Val Gly Thr Thr Ala Pro Ala
301      610          615          620
303 atg ggc gcg tgg cgc gcg acc acc ccc tcg ggc ccc tcg ccc gcc gcc      1920
304 Met Gly Ala Trp Arg Ala Thr Thr Pro Ser Gly Pro Ser Pro Ala Ala
305 625          630          635          640
307 gcc acc ccc aag gtg acc acc tac aag ccc gcc ctg ccc gcc acc gcc      1968
308 Ala Thr Pro Lys Val Thr Thr Tyr Lys Pro Ala Leu Pro Ala Thr Ala
309          645          650          655
311 aag ccc aag acc gct ggc ctc aag ctg gcc ggt gag gcc tcc acc acc      2016
312 Lys Pro Lys Thr Ala Gly Leu Lys Leu Ala Gly Glu Ala Ser Thr Thr
313          660          665          670
315 tcg acc tcg gag aac ggc gct gcc tcc aac ggc aac ggc aac ggt gcc      2064
316 Ser Thr Ser Glu Asn Gly Ala Ala Ser Asn Gly Asn Gly Asn Gly Ala
317          675          680          685
319 tcg gcc tcc aag acc tcg gct gcc aag ccc ctg gtc tcc gcc gcc acc      2112
320 Ser Ala Ser Lys Thr Ser Ala Ala Lys Pro Leu Val Ser Ala Ala Thr
321          690          695          700
323 cgc aag tcc gcc      2124
324 Arg Lys Ser Ala
325 705
328 <210> SEQ ID NO: 3
329 <211> LENGTH: 708
330 <212> TYPE: PRT
331 <213> ORGANISM: artificial sequence
333 <220> FEATURE:
334 <223> OTHER INFORMATION: fragment of the complete sequence of cDNA coding for the
GBSSI of
335      Chlamydomonas reinhardtii
337 <400> SEQUENCE: 3
339 Met Ala Val Ala Ser Thr Ser Arg Pro Ser Ser Ala Arg Pro Ile Val
340 1          5          10          15
343 Ile Asn Ala Ala Ser Phe Gly Val Lys Lys Thr Ala Asn Gln Leu Leu
344          20          25          30
347 Arg Glu Leu Ala Arg Gly Ser Ala Arg Lys Ser Thr Ser Arg Ser Ala
348          35          40          45
351 Val Thr Gly Ala Thr Gly Ala Thr Cys Ala Leu Asp Ile Val Met Val
352 50          55          60
355 Ala Ala Glu Val Ala Pro Trp Ser Lys Thr Gly Gly Leu Gly Asp Val
356 65          70          75          80
359 Thr Gly Gly Leu Pro Ile Glu Leu Val Lys Arg Gly His Arg Val Met
360          85          90          95
363 Thr Ile Ala Pro Arg Tyr Asp Gln Tyr Ala Asp Ala Trp Asp Thr Ser
364          100          105          110
367 Val Val Val Asp Ile Met Gly Glu Lys Val Arg Tyr Phe His Ser Ile
368          115          120          125
371 Lys Lys Gly Val His Arg Val Trp Ile Asp His Pro Trp Phe Leu Ala

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/980,771A

DATE: 04/11/2002

TIME: 12:45:15

Input Set : A:\EP.txt

Output Set: N:\CRF3\04112002\I980771A.raw